

MATH 343 / 643 Homework #2 INCOMPLETE

Professor Adam Kapelner

Due 11:59PM April 10, 2024

(this document last updated 1:29pm on Tuesday 26th March, 2024)

Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual “working out.” Feel free to “work out” with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read as much as you can online about the topics we covered.

The problems below are color coded: **green** problems are considered *easy* and marked “[easy]”; **yellow** problems are considered *intermediate* and marked “[harder]”, **red** problems are considered *difficult* and marked “[difficult]” and **purple** problems are extra credit. The *easy* problems are intended to be “giveaways” if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 7 points are given as a bonus if the homework is typed using L^AT_EX. Links to installing L^AT_EX and program for compiling L^AT_EX is found on the syllabus. You are encouraged to use **overleaf.com**. If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the “\vspace” command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using L^AT_EX, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: _____

Problem 1

This problem is about OLS estimation in regression. You can assume that

$$\mathbf{X} := [\mathbf{1}_n \mid \mathbf{x}_{.1} \mid \dots \mid \mathbf{x}_{.p}] \text{ with column indices } 0, 1, \dots, p \text{ and row indices } 1, 2, \dots, n$$

$$\mathbf{H} := \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{B} := (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

$$\hat{\mathbf{Y}} = \mathbf{H} \mathbf{Y} = \mathbf{X} \mathbf{B}$$

$$\mathbf{E} := \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{H}) \mathbf{Y}$$

where the entries of \mathbf{X} are assumed fixed and known and the entries of $\boldsymbol{\beta}$ are the unknown parameter).

- (a) [easy] When we “do inference” for the linear model, what is the parameter vector?
- (b) [easy] When we “do inference” for the linear model, what are considered the fixed and known quantities?
- (c) [easy] When we “do inference” for the linear model, what are considered the random quantities? And what is the notation for their corresponding realizations?
- (d) [easy] What is the “core assumption” in which the classic linear model inference follows?
- (e) [easy] From the core assumption, derive the distribution of \mathbf{B} .
- (f) [easy] From this result, derive the distribution of B_j .
- (g) [easy] From this result, derive the distribution of B_j standardized.

(h) [easy] from the core assumption, derive the distribution of \mathbf{Y} .

(i) [easy] From this result, derive the distribution of Y_i .

(j) [easy] From this result, derive the distribution of Y_i standardized.

(k) [easy] from the core assumption, derive the distribution of \mathbf{E} .

(l) [easy] From this result, derive the distribution of E_i .

(m) [easy] From this result, derive the distribution of E_i standardized.

(n) [easy] From the core assumption, show that $\frac{1}{\sigma^2} \boldsymbol{\mathcal{E}}^\top \boldsymbol{\mathcal{E}} \sim \chi_n^2$.

(o) [easy] Let $\mathbf{B}_1 = \mathbf{H}$ and let $\mathbf{B}_2 = \mathbf{I}_n - \mathbf{H}$. Justify the use of Cochran's theorem and then find the distributions of $\frac{1}{\sigma^2} \boldsymbol{\mathcal{E}}^\top \mathbf{B}_1 \boldsymbol{\mathcal{E}}$ and $\frac{1}{\sigma^2} \boldsymbol{\mathcal{E}}^\top \mathbf{B}_2 \boldsymbol{\mathcal{E}}$.

(p) [easy] Show that $\frac{1}{\sigma^2} \boldsymbol{\mathcal{E}}^\top \mathbf{B}_1 \boldsymbol{\mathcal{E}} = \frac{1}{\sigma^2} \|\mathbf{X}(\mathbf{B} - \boldsymbol{\beta})\|^2$.

(q) [harder] Why is the term $\|\mathbf{X}(\mathbf{B} - \boldsymbol{\beta})\|^2$ used to measure the model's "estimation error"?

(r) [easy] Show that $\frac{1}{\sigma^2} \boldsymbol{\mathcal{E}}^\top \mathbf{B}_2 \boldsymbol{\mathcal{E}} = \frac{1}{\sigma^2} \|\mathbf{E}\|^2$.

(s) [harder] In what scenarios is $\boldsymbol{\mathcal{E}}^\top \mathbf{B}_1 \boldsymbol{\mathcal{E}} > \boldsymbol{\mathcal{E}}^\top \mathbf{B}_2 \boldsymbol{\mathcal{E}}$?

(t) [harder] Draw an illustration of $\boldsymbol{\mathcal{E}}$ being orthogonally projected onto $\text{colsp}[\mathbf{X}]$ via projection matrix \mathbf{H} . Use the previous answers to denote the quantities of the projection and the error of the projection.

(u) [difficult] A good linear model has a large or a small projection of the error? Discuss.

(v) [easy] Find $\mathbb{E} \left[\frac{1}{\sigma^2} \|\mathbf{E}\|^2 \right]$.

(w) [easy] Show that $\frac{\|\mathbf{E}\|^2}{n-(p+1)}$ is an unbiased estimate of σ^2 .

(x) [easy] Prove that $\frac{\sqrt{n-(p+1)}(B_j - \beta_j)}{\|\mathbf{E}\| \sqrt{(\mathbf{X}^T \mathbf{X})_{j,j}^{-1}}} \sim T_{n-(p+1)}$.

- (y) [easy] Let $H_0 : \beta_j = 0$. Find the test statistic using the fact from the previous question. Let s_e denote $RMSE := \sqrt{MSE} := \sqrt{SSE/(n - (p + 1))} = \sqrt{\|e\|^2 / (n - (p + 1))}$.

- (z) [easy] Consider a new parameter of interest $\mu_\star := \mathbb{E}[Y_\star] = \mathbf{x}_\star \boldsymbol{\beta}$, this is the expected response for a unit with measurements given in row vector \mathbf{x}_\star whose first entry is 1. Prove that $\frac{\hat{Y}_\star - \mu_\star}{\sigma \sqrt{\mathbf{x}_\star (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_\star^\top}} \sim \mathcal{N}(0, 1)$.

- (aa) [easy] Prove that $\frac{\sqrt{n - (p + 1)}(\hat{Y}_\star - \mu_\star)}{\|\mathbf{E}\| \sqrt{\mathbf{x}_\star (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_\star^\top}} \sim T_{n-(p+1)}$.

- (bb) [easy] Let $H_0 : \mu_\star = 17$. Find the test statistic using the fact from the previous question. Let s_e denote the $RMSE$.

- (cc) [easy] Consider a new parameter of interest $y_\star = \mathbf{x}_\star \boldsymbol{\beta} + \epsilon_\star$, this is the response for a unit with measurements given in row vector \mathbf{x}_\star whose first entry is 1. Prove that

$$\frac{\hat{Y}_\star - y_\star}{\sigma \sqrt{1 + \mathbf{x}_\star (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_\star^\top}} \sim \mathcal{N}(0, 1).$$

- (dd) [easy] Prove that $\frac{\sqrt{n - (p + 1)}(\hat{Y}_\star - y_\star)}{\|\mathbf{E}\| \sqrt{1 + \mathbf{x}_\star (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_\star^\top}} \sim T_{n-(p+1)}.$

- (ee) [easy] Let $H_0 : y_\star = 37$. Find the test statistic using the fact from the previous question. Let s_e denote the *RMSE*.

- (ff) [difficult] Let $S \subseteq \{1, 2, \dots, p\}$, let $k := |S|$ and let $A = \{0\} \cup S^C$, its complement with zero for the index of the intercept. For convenience, assume you rearrange the columns of the design matrix so that $\mathbf{X} = [\mathbf{X}_A \mid \mathbf{X}_S]$ and the first column is $\mathbf{1}_n$. Let $\mathbf{H}_A := \mathbf{X}_A (\mathbf{X}_A^\top \mathbf{X}_A)^{-1} \mathbf{X}_A^\top$. It is obvious that $\mathbf{H} - \mathbf{H}_A$ is symmetric as both \mathbf{H} and

\mathbf{H}_A are symmetric. To prove that $\mathbf{H} - \mathbf{H}_A$ is an orthogonal projection matrix, prove that it is idempotent. Hint: use the Gram-Schmidt decomposition for both matrices and use block matrix format for \mathbf{H} .

(gg) [easy] Let $\hat{\mathbf{Y}}_A := \mathbf{H}_A \mathbf{Y}$, the orthogonal projection onto $\text{colsp}[\mathbf{X}_A]$. Prove that

$$\frac{(n - (p + 1)) \left\| \hat{\mathbf{Y}} - \hat{\mathbf{Y}}_A \right\|^2}{k \left\| \mathbf{E} \right\|^2} \sim F_{k, n-(p+1)}.$$

(hh) [difficult] Let $\hat{\mathbf{E}}_A := (\mathbf{I}_n - \mathbf{H}_A) \mathbf{Y}$, the orthogonal projection onto the $\text{colsp}[\mathbf{X}_{A^\perp}]$. Prove that $\left\| \hat{\mathbf{E}}_A \right\|^2 - \left\| \hat{\mathbf{E}} \right\|^2 = \left\| \hat{\mathbf{Y}} - \hat{\mathbf{Y}}_A \right\|^2$.

(ii) [easy] Combining the two previous problems, write the test statistic for $H_0 : \beta_S = \mathbf{0}_k$ where β_S denotes the subvector of β with indices S . Use the notation $\Delta SSE := SSE_A - SSE$ and MSE .

(jj) [difficult] Prove that the square root of the test statistic in (ii) is the same as t-test statistic from (y) when $k = 1$.

(kk) [harder] The point of this exercise is to demonstrate that the estimator used for the omnibus / global / overall F-test is nothing but a special case of the main result from (gg). Let $S = \{1, 2, \dots, p\}$ and thus $k = p$ and $A = \{0\}$. Using the result from (gg),

show that
$$\frac{(n - (p + 1)) \left\| \hat{\mathbf{Y}} - \bar{y} \mathbf{1}_n \right\|^2}{p \left\| \mathbf{E} \right\|^2} \sim F_{p, n - (p + 1)}.$$

(ll) [easy] Prove that the omnibus / global / overall F-test statistic is $\hat{F} = MSR/MSE$ by using the result from (kk).

(mm) [difficult] [MA] Prove that the distribution that realizes the R^2 metric (the proportion of response variance explained by the model) is distributed as Beta $\left(\frac{p}{2}, \frac{n-(p+1)}{2}\right)$. This amounts to proving a fact found at the bottom of the F distribution's Wikipedia page .

(nn) [easy] Prove that the maximum likelihood estimate for β is \mathbf{b} , the OLS estimator.

(oo) [harder] Prove that the maximum likelihood estimate for σ^2 is SSE/n .

(pp) [harder] Find the bias of the maximum likelihood estimator for σ^2 using your answers from (w) and (oo).

Problem 2

This problem is about two types of Bayesian estimation of the slope parameters in linear regression which lead to the ridge and lasso estimates.